UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

ON THE JOHNSON-NEYMAN TECHNIQUE AND SOME EXTENSIONS THEREOF

by

Richard F. Potthoff

January 1963

Contract No. AF 49(638)-213

The Johnson-Neyman technique is a statistical tool used most frequently in educational and psychological applications. This paper starts by briefly reviewing the Johnson-Neyman technique and suggesting when it should and should not be used; then several different modifications and extensions of the Johnson-Neyman technique, all of them conceptually simple, are proposed. The close relation between confidence intervals and regions of significance of the Johnson-Neyman type is pointed out. The problem of what to do when more than two groups are being compared is considered. The situation of more than one criterion variable is also considered.

This research was supported in part by Educational Testing Service, and in part by the Mathematics Division of the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 351

ON THE JOHNSON-NEYMAN TECHNIQUE AND SOME EXTENSIONS THEREOF[1]

by

Richard F. Potthoff
University of North Carolina

= =

1. SUMMARY AND INTRODUCTION

When measurements on one criterion variable and two or more predictor variables are available for the members of two groups (groups of classes or of individuals, e.g.), the Johnson-Neyman technique [5, 4, 1] may be used to obtain a region of values of the predictor variables for which the null hypothesis of no difference in expected criterion score between the two groups would be rejected at the .05 level (or at the α level, if we want to be more general). Thus, for any specific point P within this region of significance determined by the Johnson-Neyman technique, the number 0 does not lie within the 95% confidence interval for the difference in expected criterion score between the two groups at point P, and hence one can be at least 95% confident that there is a true difference between the two groups at point P; however, it does not follow that one can state with 95% confidence that there is a non-zero difference between the two groups simultaneously for all points in the region. A statement of this latter type is related to simultaneous confidence bounds: if the experimenter wants a region about which this sort of statement can be made, then a modification of the Johnson-Neyman technique is required, as will be described in Section 3.

[1]This research was supported in part by Educational Testing Service, and in part by the Mathematics Division of the Air Force Office of Scientific Research.

An approach which is essentially similar to the region approach, but which uses simultaneous confidence intervals in lieu of getting a region plotted, is discussed in Section 4.

The Johnson-Neyman technique was designed for a situation in which there are just two groups being compared. To handle the case where there are more than two groups, Section 5 develops a simple technique which is based on simultaneous confidence bounds for the differences between all possible pairs of groups.

Another extension of the Johnson-Neyman technique is dealt with in Section 6, where the case of more than one criterion variable is considered. Again the approach is simply that of simultaneous confidence bounds, but different tools are used this time.

The techniques suggested in this paper all require the calculation of many of the same quantities which must be computed for the Johnson-Neyman technique; beyond this point, the remaining calculations required are of an elementary nature in all of the techniques.

To start off, Section 2 will briefly review the foundations of the Johnson-Neyman technique itself, and will indicate how the technique should and should not be used.

2. THE JOHNSON-NEYMAN TECHNIQUE

We suppose that we have two groups of classes or of individuals, with n_1 members in the first group and n_2 members in the second group. Each member of each group is measured on each of r predictor (control) variables $X_1, X_2, \ldots, X_r$: let X_{ijk} denote the measurement for the k-th member of the j-th group on the i-th predictor variable ($k = 1, 2, \ldots, n_j$; $j = 1, 2$; $i = 1, 2, \ldots, r$). Each member of each group is also measured on the single criterion variable, and we will let Y_{jk} denote this criterion measurement for the k-th

member of the j-th group. (If the members of the groups are classes rather than individual students, then the Y_{jk}'s and the X_{ijk}'s would of course be class average scores rather than individual student scores). In the model we assume that the Y_{jk}'s are normally, independently, and homoscedastically distributed, and that the expectation of Y_{jk} for given X's is of the form

$$(2.1) \qquad E(Y_{jk}) = a_j + b_{1j} X_{1jk} + b_{2j} X_{2jk} + \ldots + b_{rj} X_{rjk}$$

$$(k = 1, 2, \ldots, n_j;\ j = 1, 2) \quad ,$$

where the a_j's and the b_{ij}'s are unknown regression parameters. [Some kind of attempt should be made to check that the homoscedasticity assumption is at least roughly satisfied. Two ways in which unequal variances could easily arise are (a) the variances might be different for the two groups, i.e., the Y_{1k}'s might have a variance different from the Y_{2k}'s; and (b) if the different k-values represent different classes of students, then radical differences in the class sizes might cause appreciable heteroscedasticity.7

In the original paper [5], the Johnson-Neyman technique was formulated for $r = 2$ predictor variables. However, it has been extended explicitly (see Abelson [1], e.g.) to cover the case where there is a general number (r) of predictor variables, as in the model (2.1). The essence of this development of the Johnson-Neyman technique for general r may be summarized as follows. Let C_j $(j = 1, 2)$ be an rxr matrix whose general element is

$$(2.2) \qquad c_{j,iI} = \sum_{k=1}^{n_j} X_{ijk} X_{Ijk} - \frac{\left(\sum_{k=1}^{n_j} X_{ijk}\right)\left(\sum_{k=1}^{n_j} X_{Ijk}\right)}{n_j} \quad .$$

Let $u_j (j = 1, 2)$ be an rx1 vector whose i-th element is

$$(2.3) \qquad u_{ij} = \sum_{k=1}^{n_j} X_{ijk} Y_{jk} - \frac{\left(\sum_{k=1}^{n_j} X_{ijk}\right)\left(\sum_{k=1}^{n_j} Y_{jk}\right)}{n_j} .$$

Then we write the rx1 vector

$$(2.4) \qquad \hat{b}_j = (\hat{b}_{1j}, \hat{b}_{2j}, \ldots, \hat{b}_{rj})' = C_j^{-1} u_j$$

for the estimates of the b-coefficients for $j = 1, 2$. The error sum of squares is

$$(2.5) \qquad S_e^2 = \sum_{j=1}^{2} \sum_{k=1}^{n_j} Y_{jk}^2 - \sum_{j=1}^{2} \frac{\left(\sum_{k=1}^{n_j} Y_{jk}\right)^2}{n_j} - \sum_{j=1}^{2} \sum_{i=1}^{r} \hat{b}_{ij} u_{ij} ,$$

and the error mean square is $s_e^2 = S_e^2/f$, where

$$(2.6) \qquad f = \sum_{j=1}^{2} (n_j - r - 1) .$$

Let $\bar{X}_j$ $(j = 1,2)$ denote an rx1 vector whose i-th element is

$$(2.7) \qquad \bar{X}_{ij} = \frac{\sum_{k=1}^{n_j} X_{ijk}}{n_j} .$$

Also we define

$$(2.8) \qquad \bar{Y}_j = \frac{\sum_{k=1}^{n_j} Y_{jk}}{n_j} .$$

Now let the rx1 vector

$$(2.9) \qquad X = (X_1, X_2, \ldots, X_r)'$$

denote a set of values of the control variables. We would reject at the α level the hypothesis of no difference in expected criterion score between the two groups at the point X (2.9) if

$$(2.10) \qquad \left| \frac{(\bar{Y}_2 - \hat{b}_2' \bar{X}_2) - (\bar{Y}_1 - \hat{b}_1' \bar{X}_1) + (\hat{b}_2' - \hat{b}_1') X}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2} + (X' - \bar{X}_1') C_1^{-1}(X-\bar{X}_1) + (X'-\bar{X}_2')C_2^{-1}(X-\bar{X}_2)}\sqrt{s_e^2}} \right| > t_{f;\alpha} ,$$

where $t_{f;\alpha}$ is the $(1 - \frac{\alpha}{2})$ fractile of the t distribution with f(2.6) degrees of freedom. The inequality (2.10) defines a region (which we will call R) in the r-dimensional space of the predictor variables; it is this region which is the Johnson-Neyman region of significance. We may re-write (2.10) in the form

$$(2.11) \quad \left[(\bar{Y}_2 - \hat{b}_2' \bar{X}_2 - \bar{Y}_1 + \hat{b}_1' \bar{X}_1) + (\hat{b}_2' - \hat{b}_1') X \right]^2$$

$$- t_{f;\alpha}^2 s_e^2 \left[\frac{1}{n_1} + \frac{1}{n_2} + (X'-\bar{X}_1')C_1^{-1}(X-\bar{X}_1) + (X'-\bar{X}_2')C_2^{-1}(X-\bar{X}_2) \right] > 0 ;$$

actually, there are still other ways of writing the inequality which defines R (see [17], e.g.).

The Johnson-Neyman region of significance can be plotted when $r = 2$, but the situation becomes much more troublesome when $r > 2$. Because of this difficulty in plotting R when $r > 2$, the usefulness of the Johnson-Neyman technique for general r has been questioned.

There is a second possible trouble point regarding the Johnson-Neyman technique which has perhaps never been sufficiently emphasized. As indicated in Section 1, we can be at least 95% confident (for $\alpha = .05$) that the two groups are different for any specified <u>individual</u> point in R, but we cannot be 95%

confident that the two groups are different simultaneously for all points in R. In some cases one might desire a statement of the former kind, in other cases a statement of the latter kind. For instance, if one is interested only in a single specific class of students (and if the two groups, $j = 1$ and 2, represent two curriculums, e.g.), then it would be legitimate to inquire whether the point X (the set of control variable scores) for the class falls within the region R, in order to determine whether the two curriculums may be assumed to have different effects for that class; however, for a purpose of this sort, it would appear to be not only less difficult but also more informative to obtain a confidence interval for the true difference in effects between the two curriculums for that class, rather than to undertake the labor of plotting R. On the other hand, if one is interested in making a 95% confidence statement about an entire point-set in the X-space (as would likely be the case if one were interested in over-all educational policy for many classes, rather than in a single class), then it would not be proper to use the Johnson-Neyman region R since the simultaneous confidence coefficient for such a region would generally be under 95%.

All of this suggests that perhaps some small modification or extension of the idea behind the Johnson-Neyman technique might be appropriate. In the next two sections we will suggest two possible approaches along these lines. First, Section 3 will indicate one technique for obtaining a region (point-set) whose simultaneous confidence coefficient is $\geq$ 95%. Second, Section 4 will present a confidence interval approach which, in a certain sense furnishes all the information which the region approach furnishes plus some more but which does not provide us with a region to be plotted; actually, a region approach, if interpreted in a certain way, is practically equivalent to a group of confidence statements anyway.

3. A "SIMULTANEOUS" REGION

In this section we will present one technique for obtaining what we will call a "simultaneous" region of significance, i.e., a region (which we will call R') such that, with confidence $\geq 95\%$, we can state that the two groups ($j = 1$ and 2) are different simultaneously for all points in R'. In other words, in the long run, not more than 5% of such regions R' which are calculated will contain any points at all for which the two groups are equal in expected criterion score.

Probably a number of different techniques for obtaining a simultaneous region could be devised. Computationally, the technique to be presented here bears an extremely close resemblance to the Johnson-Neyman technique, and for this reason is probably computationally simpler than most alternative techniques would be. On the other hand, it is possible that some alternative technique might produce a region which would be "larger" (in some sense) and therefore better in that respect.

Our approach will utilize a simple argument based on the simultaneous confidence bound method developed by Roy and Bose [7, Section 2.1, formula (2.1.6) especially7. Let us define

$$(3.1) \qquad d_{21} = d_{21}(X) = (a_2 - a_1) + (b_2' - b_1')X \quad ,$$

where b_j ($j = 1, 2$) is an $r \times 1$ vector containing the r b_{ij}'s. We may write

$$(3.2) \qquad \hat{d}_{21} = \hat{d}_{21}(X) = (\hat{a}_2 - \hat{a}_1) + (\hat{b}_2' - \hat{b}_1')X \quad ,$$

where

$$(3.3) \qquad \hat{a}_j = \bar{Y}_j - \hat{b}_j' \bar{X}_j$$

for $j = 1, 2$. Also we define

$$(3.4) \qquad v_j = v_j(X) = \frac{1}{n_j} + (X' - \bar{X}_j') C_j^{-1} (X - \bar{X}_j)$$

for $j = 1, 2$. Now we are ready to utilize [7, formula (2.1.6)] to establish that $100(1-\alpha)\%$ simultaneous confidence intervals for the functions $d_{21}(X)$ for all possible points X in the r-dimensional X-space are given by

$$(3.5) \qquad \hat{d}_{21}(X) \pm \sqrt{(r+1)F_{r+1,f;\alpha}} \sqrt{v_2(X) + v_1(X)} \sqrt{s_e^2} \quad ,$$

where $F_{r+1,f;\alpha}$ denotes the $(1-\alpha)$ fractile of the F distribution with $r+1$ and f degrees of freedom. If we examine (3.5), we can conclude that, with simultaneous confidence coefficient $\geq 100(1-\alpha)\%$, we can state that $d_{21}(X)$ is $\neq 0$ (i.e., that the two groups are different) for all points X such that

$$(3.6) \qquad [\hat{d}_{21}(X)]^2 > (r+1) F_{r+1,f;\alpha} [v_2(X) + v_1(X)] s_e^2 \quad .$$

Thus (3.6) defines the simultaneous region R'.

Note that the defining inequality for the simultaneous region R' (3.6) is basically identical with the defining inequality for the Johnson-Neyman region R (2.11), with the one exception that $t_{f;\alpha}^2$ in (2.11) is replaced by $(r+1)F_{r+1,f;\alpha}$ in (3.6). Hence the computational procedure for obtaining R' is essentially the same as that for obtaining the Johnson-Neyman region R.

As is the case with the Johnson-Neyman region, the region R' will consist of two parts: one part where the statement is made that $d_{21}(X) > 0$, and another part containing those X's for which we state that $d_{21}(X) < 0$. Recognizing

these two separate parts is thus more explicit than merely stating that $d_{21}(X) \neq 0$ throughout R'. As with the Johnson-Neyman region, one or both of these two parts of R' may in some cases be null (i.e., contain no points at all).

It appears that the "simultaneous" region R' (3.6) will always be a (proper) subset of the Johnson-Neyman region R (2.11) (i.e., every point in R' will also be in R, but not every point in R will be in R'). Thus R' could turn out to be "smaller" than we might like it to be. However, it has been pointed out (see Scheffé [8, p. 71], e.g.) that, when we are dealing with simultaneous confidence bounds, it may be most sensible to choose an α value somewhat larger than what we would customarily choose for simple confidence bounds. Hence the choice of a larger-than-usual α would result in R' not being so "small".

As already indicated, the region R' (3.6) may not be the most desirable $100\,(1-\alpha)\%$ simultaneous region which can be found: if we examine more closely the situation for $r = 1$, we will suspect that, for general r, it should somehow be possible for the region R' (3.6) to be improved upon (although such improvement might well be at the expense of increased computational difficulty). For $r = 1$, if we write $X_{10} = -(a_2 - a_1)/(b_{12} - b_{11})$ for the value of X_1 for which $d_{21}(X)$ [i.e., $d_{21}(X_1)$] is 0 (we assume $b_{12} \neq b_{11}$), then a $100(1-\alpha)\%$ confidence region for X_{10} (see Fisher [3, pp. 144-146], e.g.) is specified by the inequality

$$(3.7) \quad [\hat{d}_{21}(X_{10})]^2 \leq t^2_{f;\alpha}\,[v_2(X_{10}) + v_1(X_{10})]\,s_e^2 \quad .$$

If we state that we are $100\,(1-\alpha)\%$ confident that X_{10} lies in the region (3.7), this is equivalent to stating that we are $100\,(1-\alpha)\%$ confident that

X_{10} does not lie in the complement of (3.7), which (in turn) is the same thing as saying that we are $100(1-\alpha)\%$ confident that $d_{21}(X) \neq 0$ for all X such that

$$(3.8) \qquad [\hat{d}_{21}(X)]^2 > t^2_{f;\alpha} [v_2(X) + v_1(X)] s_e^2 \quad .$$

But we note that (3.8) is the same as (3.6), except that instead of $2F_{2,f;\alpha}$ in (3.6) we have the apparently smaller quantity $F_{1,f;\alpha}$ in (3.8). Hence R' (3.6) can evidently be improved upon for r=1, and although unfortunately no ready generalization of (3.8) for r>1 [such as an inequality like (3.8) but with $t^2_{f;\alpha}$ replaced by $r F_{r,f;\alpha}$ for general r, e.g.] could be discovered, one would still suspect from this that R' (3.6) could be improved upon for $r > 1$ also. [We take note finally, however, of a curious fact. For the region (3.6), we can state with $\geq 100(1-\alpha)\%$ confidence not only that $d_{21}(X) \neq 0$ for all X satisfying (3.6), but more specifically that $d_{21}(X) > 0$ for all X satisfying (3.6) for which $\hat{d}_{21}(X) > 0$ and that $d_{21}(X) < 0$ for all X satisfying (3.6) for which $\hat{d}_{21}(X) < 0$. For the region (3.8), however, we apparently can state with $100(1-\alpha)\%$ confidence only that $d_{21}(X) \neq 0$ for all X satisfying (3.8), and we cannot make any more specific statement about the sign of $d_{21}(X)$ without causing the confidence coefficient to drop below $100(1-\alpha)\%$. This would be a drawback of using the region (3.8).]

4. SIMULTANEOUS CONFIDENCE INTERVALS

The Johnson-Neyman approach and the simultaneous region approach which was suggested in Section 3 both require that a region be obtained and usually plotted. As already indicated, this may not be an easy task, especially when $r > 2$. Therefore it is appropriate to consider an alternative approach.

In a certain sense, the confidence interval approach which we are about to suggest is almost equivalent to the region approach anyway; for both the region R (2.11) and the region R' (3.6) are essentially based on confidence bounds to begin with (the former on simple confidence bounds, the latter on simultaneous confidence bounds). The confidence bounds associated with the region R (2.11) are simple confidence bounds on the function $d_{21}(X)$ for a specified X: for any individual X, we can be $100(1-\alpha)^{0}/_{0}$ confident that $d_{21}(X)$ lies within the interval

$$(4.1) \qquad \hat{d}_{21}(X) \pm t_{f;\alpha} \sqrt{v_2(X) + v_1(X)} \sqrt{s_e^2} \quad .$$

The relation between the Johnson-Neyman region R (2.11) and the confidence interval formula (4.1) is simply this: the number 0 will fall outside of the interval (4.1) if and only if X lies in the region R (2.11) .

The confidence bounds associated with the region R' (3.6) were actually used as a step in the argument which developed (3.6), and are specified by the simultaneous confidence interval formula (3.5). The number 0 will fall outside of the interval (3.5) if and only if X lies in the region R' (3.6).

What we will now propose is that, in some situations, it may be more sensible to utilize the confidence intervals (4.1) or (3.5) in lieu of obtaining the regions (2.11) or (3.6) respectively. We have already noted the close relation between the confidence intervals and the regions. In one respect, the confidence interval approach actually gives more information than the region approach: the former furnishes us with a specific confidence interval for $d_{21}(X)$ for any X, whereas the latter does not provide us with the interval but rather tells us only (for every X) whether or not this confidence interval contains the value 0. On the other hand, though, the confidence interval approach does not furnish us

explicitly wit· .ny graphical "region of significance".

The simp·e confidence interval (4.1), which is associated with the Johnson-Neyman region R (2.11), may be used (as was already indicated in Section 2) when we want confidence bounds on $d_{21}(X)$ for a single specific X. The use of the confidence interval approach with respect to the simultaneous bounds (3.5), however, requires somewhat more discussion than the formula (4.1) because there are different ways in which (3.5) can be utilized; we mention some possibilities:

(a) The inte.vals (3.5) could be calculated for a certain number of strategically - placed points in the X-space. For example, a certain number of equally-spaced values could be chosen for each one of the r control variables, and (3.5) could then be calculated for all possible combinations of these values, thereby furnishing us with the intervals (3.5) for all the points on a sort of r-dimensional grid in the X-space.

(b) We might learn something interesting by calculating (3.5) for the $(n_1 + n_2)$ points in the X-space represented by the $(X_{11k}, X_{21k}, \ldots, X_{r1k})$'s $(k = 1,2,\ldots, n_1)$ and the $(X_{12k}, X_{22k}, \ldots, X_{r2k})$'s $(k = 1,2,\ldots,n_2)$. The idea would be that these $(n_1 + n_2)$ points would in many cases be a reasonable cross-section of the particular population of points in the X-space in which we would be interested. Furthermore, we might classify each of the $(n_1 + n_2)$ intervals into one of three groups: intervals lying wholly above 0, intervals lying wholly below 0, and intervals which contain 0 [and the last group could be split into two parts, if desired, according to whether $\hat{d}_{21}(X) \geq 0$ or < 0]. It might then be instructive to tally the number of intervals in each group.

(c) We might simply have a specific list of points in the X-space for which we wish to obtain simultaneous confidence intervals, where the points might (e.g.) be the X-scores for a list of certain classes (of students) which we are interested

in and which need not be among the $(n_1 + n_2)$ classes appearing in the data.

5. THE CASE OF MORE THAN TWO GROUPS

The methods we have discussed so far have been for situations where there are only two groups ($j = 1$ and 2). These two groups may represent (e.g.) two different curriculums or two different teaching techniques. We now consider a more general situation where there are g groups ($j = 1,2,\ldots,g$) being compared rather than just two. We will use the same notation and assumptions as before [see particularly (2.1 - 2.4), (2.7 - 2.8), (3.3 - 3.4)], but with the understanding that the subscript j may assume the values $1,2,\ldots,g$ instead of just the values 1 and 2.

A total of $\frac{1}{2}g(g-1)$ different possible pairs of groups can be chosen from among the g groups. Let (j, J) be any such pair (we assume $J > j$ for definiteness). Then, generalizing (3.1), we can define

$$(5.1) \qquad d_{Jj} = d_{Jj}(X) = (a_J - a_j) + (b'_J - b'_j)\,X \qquad (J > j) .$$

Thus, for the g-group situation, we are potentially interested in the $\frac{1}{2}g(g-1)$ differences d_{Jj} (5.1) rather than just in a single difference d_{21} (3.1).

Our problem is to generalize the techniques available for the two-group situation in order to obtain ways to handle the g-group situation. The idea suggested here will be a simple one which again will be based on the simultaneous confidence bound method given by Roy and Bose [7, Section 2.1]: we will obtain simultaneous confidence bounds on all $\frac{1}{2}g(g-1)$ possible d_{Jj}'s .

If we are interested in a single specific X, then $100(1-\alpha)\%$ simultaneous confidence intervals for $d_{Jj}(X)$ <u>for all possible (j,J)</u> but for this single X are given by

$$(5.2) \qquad \hat{d}_{Jj}(X) \pm \sqrt{(g-1)F_{g-1,f;\alpha}}\ \sqrt{v_J(X) + v_j(X)}\ \sqrt{\frac{s_e^2}{f}} \quad ,$$

where $\hat{d}_{Jj}(X)$ is defined analogously to (3.2), and where S_e^2 and f are defined the same as (2.5) and (2.6) respectively except with $\sum_{j=1}^{2}$ replaced wherever it appears by $\sum_{j=1}^{g}$. If we are interested in simultaneous bounds for all X, then $100(1-\alpha)\%$ simultaneous confidence intervals for $d_{Jj}(X)$ <u>for all possible (j,J)</u> and <u>for all possible X</u> are given by

$$(5.3) \quad \hat{d}_{Jj}(X) \pm \sqrt{(r+1)(g-1)F_{(r+1)(g-1),f;\alpha}} \sqrt{v_J(X) + v_j(X)} \sqrt{\frac{S_e^2}{f}} ,$$

where S_e^2 and f are defined the same as for (5.2). The formulas (5.2) and (5.3) are both immediate consequences of [7, formula (2.1.6)]. Note that (5.2) is a generalization of (4.1), while (5.3) is a generalization of (3.5). Many of the remarks about the confidence interval approach which were made in Section 4 can also be applied (with appropriate modification) to the confidence intervals (5.2) and (5.3).

If a region approach is preferred to a confidence interval approach, it is possible to obtain regions of significance corresponding to either (5.2) or (5.3). However, if all $\frac{1}{2}g(g-1)$ pairs (j, J) were considered, it would be necessary [in the case of <u>either</u> (5.2) or (5.3)] to obtain $\frac{1}{2}g(g-1)$ different regions (i.e., one region for each comparison). Statements could be made with $100(1-\alpha)\%$ confidence with respect to all $\frac{1}{2}g(g-1)$ regions simultaneously. The regions corresponding to (5.3), for example, are based on the inequalities

$$(5.4) \quad [\hat{d}_{Jj}(X)]^2 > (r+1)(g-1)F_{(r+1)(g-1),f;\alpha} [v_J(X) + v_j(X)] \frac{S_e^2}{f} ,$$

and so (5.4) is a generalization of (3.6) .

6. THE CASE OF MORE THAN ONE CRITERION VARIABLE

So far we have considered only the case of a single criterion variable, Y_{jk}. However, situations are sometimes encountered where there are two or more criterion variables. For example, suppose that we have an experiment with three curriculums, and suppose that each curriculum has a different test associated with it which is customarily administered to the students at the end of the course. In our experiment, however, suppose that all three tests are administered at the end of the course to all classes in each of the three groups, in order to obtain all the information which might be necessary for a fair comparison. Then the three tests would constitute three different criterion variables.

Thus we are faced with the problem of extending the techniques described earlier in the paper to the case of multiple criterion variables. Our model is now a multivariate one; instead of (2.1) we have the model equation

$$(6.1) \qquad E(Y_{jk}^{(\ell)}) = a_j^{(\ell)} + b_{1j}^{(\ell)} X_{1jk} + b_{2j}^{(\ell)} X_{2jk} + \cdots + b_{rj}^{(\ell)} X_{rjk} \quad ,$$

where the superscript ℓ is the index referring to the criterion variable. We suppose that there are p criterion variables, so that $\ell = 1,2,\ldots,p$. One would ordinarily assume for each fixed (j,k) that the p $Y_{jk}^{(\ell)}$'s follow a multivariate normal distribution with unknown variance matrix $\Sigma(pxp)$, although for the particular technique which we are about to propose it so happens that it will be sufficient to assume only that the univariate marginal distributions are normal and homoscedastic for each of the p variates.

To attack the problem of the multiple criterion variables, we will start by supposing that we are basically interested in obtaining simultaneous confidence bounds on the differences

$$(6.2) \qquad d_{Jj}^{(\ell)} = d_{Jj}^{(\ell)}(X) = (a_J^{(\ell)} - a_j^{(\ell)}) + (b_J^{(\ell)'} - b_j^{(\ell)'})X$$

for all (j, J) and for all ℓ, and either for all X or for a single specific X. [In (6.2), $b_j^{(\ell)}$ is of course a vector containing the $b_{ij}^{(\ell)}$'s .]

It would be valid to apply Roy's formula for simultaneous multivariate confidence bounds (see [6, p. 101, formula (14.6.3)], e.g.) here, but some rough investigations indicate that an alternative technique probably provides shorter confidence intervals in this particular case. We are interested in confidence bounds on a type of function (6.2) which is rather specialized in the sense that any single function (6.2) involves but one value of ℓ; and Roy's general formula [6, formula (14.6.3)] appears to be most efficient not for such relatively specialized functions, but rather for functions involving parameters associated with more than one variate.

The technique of Roy and Bose [7, Section 2.1] which we employed to advantage in both Section 4 and Section 5 cannot, of course, be appealed to again here, because it applies only to a univariate situation. The device we will suggest for getting simultaneous bounds on the functions (6.2) is conceptually a simple one, and is based on essentially the same idea which was employed by Dunn [2]: we just generalize any of our previous confidence interval formulas by substituting (α/p) for α (and attaching the superscripts ℓ in the appropriate places), and we end up with simultaneous bounds which apply (simultaneously) to all p values of ℓ.

This can be done with any of the four confidence interval formulas (4.1), (3.5), (5.2), or (5.3), depending on our needs. For example, if we generalize (5.3), we determine that $100(1-\alpha)\%$ simultaneous confidence intervals for $d_{Jj}^{(\ell)}(X)$ (6.2) <u>for all possible (j, J)</u> and <u>for all ℓ</u> and <u>for all possible X</u>

are given by

$$(6.3)\quad \hat{d}^{(\ell)}_{Jj}(X) \pm \sqrt{(r+1)(g-1)F_{(r+1)(g-1),f;\frac{\alpha}{p}}}\ \sqrt{v_J(X) + v_j(X)}\ \sqrt{\frac{S_e^{2(\ell)}}{f}}\ ,$$

where $\hat{d}^{(\ell)}_{Jj}(X)$ and $S_e^{2(\ell)}$ are the same functions of the $Y^{(\ell)}_{jk}$'s as $\hat{d}_{Jj}(X)$ and S_e^2 are (respectively) of the Y_{jk}'s .

If, for the case of p criterion variables, a region approach is preferred to a confidence interval approach, this can be arranged in a manner analogous to that employed in previous sections. However, there will be p times as many regions to obtain for the p-variate case as there would be for the corresponding univariate case.

Finally, a warning note should be sounded. As our techniques are adapted to more and more complex situations (which is the way they have developed during the course of this paper), we may in some cases become faced with wider and wider confidence intervals and with smaller and smaller regions of significance. In fact, confidence intervals might turn out to be so wide as to be useless, and regions can be so small as to be useless. When this sort of thing happens, it means that the sample sizes (the n_j's) are too small to furnish adequate information. In general, it would appear that this problem of sample sizes being too small might arise more frequently with the more complicated types of confidence intervals and regions.

7. ACKNOWLEDGMENTS

Some of the problems which were considered in this paper were suggested by R. Darrell Bock of the Psychometric Laboratory, University of North Carolina, and by Frederic M. Lord of Educational Testing Service. The author profited from helpful discussions with both of these individuals.

REFERENCES

[1] Abelson, Robert P. A note on the Neyman-Johnson technique. Psychometrika, 1953, vol. 18, pp. 213-218.

[2] Dunn, Olive Jean. Multiple comparisons among means. J. Amer. Statist. Assoc., 1961, vol. 56, pp. 52-64.

[3] Fisher, R. A. Statistical Methods for Research Workers (10th edition). Edinburgh: Oliver and Boyd, 1946 (reprinted 1948).

[4] Johnson, Palmer O., and Fay, Leo C. The Johnson-Neyman technique, its theory and application. Psychometrika, 1950, vol. 15, pp. 349-367.

[5] Johnson, Palmer O., and Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. Statistical Research Memoirs, 1936, vol. 1, pp. 57-93.

[6] Roy, S. N. Some Aspects of Multivariate Analysis. New York: Wiley, 1957.

[7] Roy, S. N., and Bose, R. C. Simultaneous confidence interval estimation. Ann. Math. Statist., 1953, vol. 24, pp. 513-536.

[8] Scheffé, Henry. The Analysis of Variance. New York: Wiley, 1959.